





LIBRARY  
OF THE  
MASSACHUSETTS INSTITUTE  
OF TECHNOLOGY





HD28  
.M414  
no.312-  
68

WORKING PAPER  
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

A CONTINUOUS LIMIT FOR  
THE "CONTAGIOUS" BINOMIAL DISTRIBUTION

312-68

David B. Montgomery  
//

MASSACHUSETTS  
INSTITUTE OF TECHNOLOGY  
50 MEMORIAL DRIVE  
CAMBRIDGE, MASSACHUSETTS 02139



A CONTINUOUS LIMIT FOR  
THE "CONTAGIOUS" BINOMIAL DISTRIBUTION

312-68

David B. Montgomery

The author is Assistant Professor of Management in the Sloan School of Management, Massachusetts Institute of Technology, and is a member of the Affiliated Faculty of the Operations Research Center at M.I.T.

HD 28  
m H H  
no 212-68

RECEIVED  
MAY 8 1968  
M. I. T. LIBRARIES



# A Continuous Limit for the "Contagious" Binomial Distribution

David B. Montgomery  
Massachusetts Institute of Technology

## 1. Introduction

The "contagious" binomial distribution is a discrete probability distribution which arises as the steady-state distribution of the system state variable in certain stochastic response models.<sup>1</sup> Social scientists who have used these models have expressed a need for a limiting form of this distribution, but have been unable to derive it.<sup>2</sup> The purpose of this paper is to show that the desired limit for the "contagious" binomial is the beta distribution.

In this paper we first present the general response model which has the "contagious" binomial as its steady-state distribution. We then develop the beta limit. Finally, we examine two major types of application of the general model.

## 2. The General Model

In the general formulation the basic model component will be referred to as an element. The total behavior of the element in the model will be termed system behavior. In the discussion of the applications of the model presented below, the terms element and system will be identified with the nomenclature which has been used in the application of this model type to particular stochastic response situations.



Discussion will be facilitated by first considering the notation which shall be used. In the general model an element will be uniquely associated with one of two states, A or B. The remainder of the notation is as follows:

$N$  - the total number of elements in the system.

$i$  - the number of elements out of the total of  $N$  elements that are in state A at any particular time. Thus  $i$  is a random variable which represents the state of the system. Clearly,  $0 \leq i \leq N$ .

$\alpha$  - the inherent transition intensity<sup>3</sup>, or propensity for an element in state B to shift to state A.

$\beta$  - the inherent transition intensity or propensity for an element in state A to shift to state B.

$\gamma$  - the additive incremental influence on the transition intensity of an element toward the opposite state exerted by each element in the opposite state.

Thus the propensity for an element in state B to shift to state A is composed of two parts: an inherent propensity,  $\alpha$ , and an additive attractive influence of the  $i$  elements already in state A,  $i\gamma$ . Similarly the propensity for an element to shift from state A to state B is  $\beta + (N-i)\gamma$ , since there are  $N-i$  elements currently associated with state B.

At the level of the elements the process has two states. That is, each element is in either state A or state B. But at the level of the system there are  $N + 1$  states corresponding to the  $N + 1$  possible values of the system state variable,  $i$ . The model gives rise to the following system of differential equations on the system state probabilities at time  $t$  which are denoted by  $p_i(t)$ :



$$\frac{dp_0(t)}{dt} = -N\alpha p_0(t) + [\beta + (N-1)\gamma] p_1(t) \quad \text{for } i = 0 \quad (1)$$

$$\begin{aligned} \frac{dp_i(t)}{dt} = & -[(N-1)(\alpha + i\gamma) + i[\beta + (N-1)\gamma]] p_i(t) \\ & + (N-i+1)[\alpha + (i-1)\gamma] p_{i-1}(t) \\ & + (i+1)[\beta + (N-i-1)\gamma] p_{i+1}(t) \quad \text{for } 0 < i < N \end{aligned}$$

$$\frac{dp_N(t)}{dt} = -N\beta p_N(t) + [\alpha + (N-1)\gamma] p_{N-1}(t) \quad \text{for } i = N$$

The steady-state distribution for the state of the system may be found by the simultaneous solution of the  $N+1$  equations given as (1) when

$$\frac{dp_i(t)}{dt} = 0 \quad \text{for } i = 0, 1, \dots, N \quad \text{and using the fact that} \quad \sum_{i=0}^N p_i(t) = 1.$$

Coleman<sup>4</sup> presents this steady-state distribution as

$$p_i = \binom{N}{i} \frac{\Gamma(\frac{\alpha}{\gamma} + i) \Gamma(N + \frac{\beta}{\gamma} - i) \Gamma(\frac{\alpha + \beta}{\gamma})}{\Gamma(\frac{\alpha}{\gamma}) \Gamma(N + \frac{\alpha + \beta}{\gamma}) \Gamma(\frac{\beta}{\gamma})} \quad (2)$$

where  $\binom{N}{i} = \frac{N!}{i! (N-i)!}$

Thus, at stochastic equilibrium the distribution of  $i$ , the number of elements in state A, is given by (2). The concern in this paper is to derive the distribution of  $X = i/N$  as  $N$  goes to infinity. That is, the distribution of the proportion of elements in state A is sought as the number of elements goes to infinity.



### 3. The Beta Limit

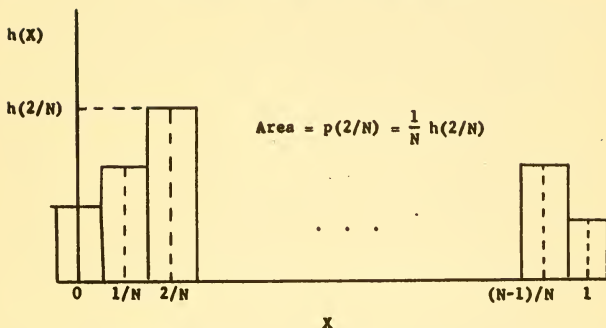
In order to determine the function which will yield a continuous probability density function for  $X$ , it is first necessary to consider what occurs as  $N$  goes to infinity in such a way that  $X = i/N$  remains constant. The latter restriction is included to ensure that the cumulative probability  $p[X \leq C]$ , where  $C$  is some constant between zero and one, remains constant as  $N$  goes to infinity.<sup>5</sup>

For  $N$  finite, the discrete mass function  $p(X = i/N)$  for  $X$  may be represented by a histogram such as that presented in Figure 1 where  $h(X)$  denotes the height of the histogram at  $X$ .





FIGURE 1  
HISTOGRAM OF THE MASS FUNCTION  $p(X)$



Note that the width of each interval is  $1/N$  and the area of the rectangle about  $X$  is the probability that the system has the proportion  $X = i/N$  of its  $N$  elements associated with response A. That is,

$$p(X) = \frac{1}{N} h(X).$$

Hence

$$h(X) = Np(X).$$

Suppose now that in the interval  $(0,1)$  more and more elements are packed in, that is,  $N$  becomes large. In this case  $1/N$  will become very small, and in the limit the  $h(X)$  will be so close together that they will trace out a continuous curve. This continuous curve is the continuous p.d.f. of  $X = i/N$  when  $N \rightarrow \infty$  in such a way that  $X = i/N$  remains constant. That is, where  $C$  indicates that  $X = i/N$  remains constant,

$$\begin{array}{lll} \lim_{\substack{N \rightarrow \infty \\ X=(i/N)=C}} h(X) & = & \lim_{\substack{N \rightarrow \infty \\ X=(i/N)=C}} N p(X) = f(X) \end{array} \quad (3)$$



Thus (3) gives the function which must be taken to the limit in order to obtain the limiting probability density function of  $X$ ,  $f(X)$ .

For any fixed  $N$ , there is a direct equivalence between  $p(X)$  in (3) and the  $p_i$  given by (2). Clearly, for some fixed  $N$ , the event  $X = i/N$  occurs if, and only if, the event  $i$  occurs. Hence for fixed  $N$  the events  $X = i/N$  and  $i$  are equivalent.

Thus  $p(X=i/N) = p_i$ . But the steady-state distribution of  $i$  for some fixed  $N$  has been given in (2). Hence (3) may be written as

$$f(X) = \lim_{\substack{N \rightarrow \infty \\ i/N=C}} N p_i \quad (4)$$

$$= \lim_{\substack{N \rightarrow \infty \\ XN=i=NC}} N \binom{N}{i} \frac{\Gamma(\frac{\alpha}{\gamma} + 1) \Gamma(N + \frac{\beta}{\gamma} - 1) \Gamma(\frac{\alpha + \beta}{\gamma})}{\Gamma(\frac{\alpha}{\gamma}) \Gamma(N + \frac{\alpha + \beta}{\gamma}) \Gamma(\frac{\beta}{\gamma})}$$

A result on the limiting behavior of gamma functions when a term in the gamma argument increases without limit is needed if the limit on the right hand side of (4) is to be found. It is known that<sup>6</sup>:

$$\lim_{\alpha \rightarrow \infty} \frac{\Gamma(\alpha + a)}{\alpha^a \Gamma(\alpha)} = 1. \quad (5)$$

$$\text{Furthermore, if } \lim_{\alpha \rightarrow \infty} \frac{A(\alpha)}{a(\alpha)} = \lim_{\alpha \rightarrow \infty} \frac{B(\alpha)}{b(\alpha)} = \lim_{\alpha \rightarrow \infty} \frac{C(\alpha)}{c(\alpha)} = \lim_{\alpha \rightarrow \infty} \frac{D(\alpha)}{d(\alpha)} = 1$$

$$\text{and if } \lim_{\alpha \rightarrow \infty} \frac{a(\alpha) b(\alpha)}{c(\alpha) d(\alpha)} = k, \text{ then}$$

$$\lim_{\alpha \rightarrow \infty} \frac{A(\alpha) B(\alpha)}{C(\alpha) D(\alpha)} = k \quad (6)$$

Using these results in (4) we have



$$\begin{aligned}
 f(X) &= \lim_{N \rightarrow \infty} \sum_{i=1}^N p_i \quad (7) \\
 &= \lim_{N \rightarrow \infty} \frac{\Gamma(\frac{\alpha + \beta}{\gamma}) N^{\frac{\alpha + \beta}{\gamma}} \Gamma(N) N^{\alpha/\gamma} X^{\alpha/\gamma} \Gamma(NX) N^{\beta/\gamma} (1-X)^{\beta/\gamma} \Gamma(N[1-X])}{\Gamma(\frac{\alpha}{\gamma}) \Gamma(\frac{\beta}{\gamma}) N^{\alpha/\gamma} X^{\alpha/\gamma} \Gamma(NX) N^{\beta/\gamma} (1-X)^{\beta/\gamma} \Gamma(N[1-X])} \\
 &= \frac{\Gamma(\frac{\alpha + \beta}{\gamma})}{\Gamma(\frac{\alpha}{\gamma}) \Gamma(\frac{\beta}{\gamma})} X^{\alpha/\gamma - 1} (1-X)^{\beta/\gamma - 1}
 \end{aligned}$$

But (7) is just the beta distribution with mean,  $\alpha/(\alpha + \beta)$ , and variance,  $\alpha\beta/(\alpha + \beta)^2 (\alpha + \beta \gamma)$ .<sup>7</sup> Thus the infinite element, continuous counterpart of the "contagious" binomial distribution is the beta distribution given in (7).

#### 4. Applications of the General Model

There have been two principal classes of application of the general model. These classes are considered below along with an indication of the motivation for seeking the limiting distribution in each case.

##### Class I. Models of Change and Response Uncertainty

The focus in this class of applications has been upon modeling the choice between two alternatives in situations where a respondent chooses A versus B with probability  $P(A)$  on any given choice occasion and where  $P(A)$  may change between successive choice occasions. Thus the model has been used in situations where choice is stochastically determined and where the probability of making a particular choice is non-stationary. This form of the general model has been applied to data from a social psychological experiment and to consumer brand choice.<sup>8</sup>



In these applications the elements of the general model are considered to be hypothetical response elements, similar in nature to the stimulus elements of stimulus sampling theory.<sup>9</sup> The system in the general model is now the individual respondent. At any time  $t$  an individual's probability of making response A versus response B is given by  $X = i/N$ , the proportion of his  $N$  response elements which are currently associated with response A -- that is, in state A. Thus in this case (2) represents the steady-state distribution of an individual respondent's probability of making response A versus response B. In an application to consumer brand choice response A might be a purchase of a brand of particular interest to us and response B a purchase of some other brand. Note that even at stochastic equilibrium, an individual's probability of making response A may change.

The desirability of letting the number of response elements increase without limit in this case is clear. A model which allows the probability that an individual will make response A to take on any value between zero and one is more satisfying than one which constrains this probability to certain discrete values. For example, a model which considers each individual to have two response elements implies that an individual's probability of making response A is either 0,  $1/2$ , or 1. A model which allows each respondent to have an unlimited number of response elements allows this probability,  $X$ , to be a continuous measure. Hence, one motivation for seeking the limit of (2) in this case is to develop the steady-state distribution for a more theoretically satisfying model of the choice probabilities.





In addition, this limiting result may be of use in estimating the distribution of  $P(A)$  across a population of respondents all having the same  $\alpha$ ,  $\beta$ , and  $\gamma$ .<sup>10</sup> The fact that the beta distribution is a well tabled distribution makes this limiting result especially useful in such cases.

## Class II Reward Models of Interpersonal Influence

Coleman has presented models of the process of interpersonal influence where the behavior of others influences an individual to exhibit similar behavior. In these applications the elements of the general model correspond to individuals and the system corresponds to the group. The parameters  $\alpha$  and  $\beta$  correspond to the inherent propensities of individuals to shift from activity D to activity A and from A to B, respectively. The parameter  $\gamma$  measures the influence which the behavior of other individuals has over the behavior of any given individual. Coleman has utilized this form of the model to analyze the impact of interpersonal influence upon voting behavior in small groups.

The motivation for developing the limiting form of (2) in this case is one of convenience rather than theoretical desirability. For moderately large groups, say groups having over twenty-five members, (2) will have  $N > 25$  and will be computationally inconvenient. The well tabled beta form will often prove to be a useful approximation in such cases.

## 5. Summary

This paper has demonstrated that the beta distribution is the limit distribution of the "contagious" binomial. The structure of certain models which have the "contagious" binomial as the steady-state distribution was illustrated.



#### FOOTNOTES

1. Coleman has coined the term "contagious binomial" to describe the distribution considered in this paper.
2. Coleman [3, p. 413] expressed the need for this limit but indicated that he had been unable to derive it.
3. For a discussion of the notion of transition intensity in relation to continuous time stochastic processes see [5, pp.423-8].
4. See [3, pg. 345].
5. This restriction is analogous to that used in the development of the Poisson limit of the binomial distribution. See [5].
6. See [4, Chapter 9].
7. See [9, pg. 216].
8. See [2], [3, Chapter 13], [7, Chapter 5], and [8].
9. For the derivation of this model type from an axiomatic system analogous to those used in stimulus sampling theory see [7, Chapter 1-3].
10. See [7, pp. 98-102] for a suggested use of the limiting distribution in this context.
11. See [1] and [3, pp. 343-353].



## REFERENCES

- [1] Coleman, James S. "Reward Structures and the Allocation of Effort". In Criswell, Solomon, and Suppes (Eds.), Mathematical Methods in Small Group Processes. Stanford, Calif.: Stanford University Press, 1962, pp. 119-32.
- [2] Coleman, James S. Models of Change and Response Uncertainty. Englewood Cliffs, N. J.: Prentice-Hall, 1964.
- [3] Coleman, James S. Introduction to Mathematical Sociology. New York: The Free Press of Glencoe, 1964.
- [4] Copson, E. T. An Introduction to the Theory of Functions of a Complex Variable. London: Oxford University Press, 1935.
- [5] Feller, William. An Introduction to Probability Theory and Its Applications. Vol. 1, Second Edition. New York: Wiley, 1957.
- [6] Kaplan, Wilfred. Advanced Calculus. Reading, Mass.: Addison-Wesley, 1952.
- [7] Montgomery, David B. "A Probability Diffusion Model of Dynamic Market Behavior". Working Paper No. 205-66, Alfred P. Sloan School of Management, M. I. T.
- [8] Montgomery, D.B. "A Stochastic Response Model with Application to Brand Choice", Working Paper No. 238-67, Alfred P. Sloan School of Management, M.I.T.
- [9] Raiffa, H. and Schlaifer, R. Applied Statistical Decision Theory. Boston, Mass.: Division of Research, Harvard Business School, 1961.





BASCOM  
Date Due

~~MAX 22 '78~~

AG 29 '91

Lib-26-67



MIT LIBRARIES DUPL



3 9080 003 702 021

304-68

MIT LIBRARIES

DUPL



3 9080 003 671 044

305-68

MIT LIBRARIES

DUPL



3 9080 003 702 112

306-68

MIT LIBRARIES

DUPL



3 9080 003 702 096

307-68

MIT LIBRARIES

DUPL



3 9080 003 670 830

308-68

MIT LIBRARIES

DUPL



3 9080 003 670 855

309-68

MIT LIBRARIES

DUPL



3 9080 003 671 119

310-68

MIT LIBRARIES

DUPL



3 9080 003 671 093

311-68

MIT LIBRARIES

DUPL



3 9080 003 671 069

312-68

MIT LIBRARIES

DUPL



3 9080 003 671 085

313-68

